ISSN: 2472-3703 DOI: 10.6911/WSRJ.202611_12(11).0007

Research on Surface Defect Detection of Photovoltaic Panels Based on an Improved YOLOv8n

Rui Wang, Heng Xiao, Jiaqi Tang, Huang Wang, Jixue Xiao*

Xihua University, Chengdu, Sichuan 610039, China

* Corresponding author

Abstract

Photovoltaic panels exposed to outdoor environments for long periods may develop defects such as cracks, micro-cracks, and short-circuited cells, which degrade their electrical performance and operational reliability. Electroluminescence (EL) imaging reveals internal structural differences of solar cells and provides an effective basis for visual defect inspection. However, micro-cracks in EL images are typically thin and of low local contrast, and the scale difference among defect categories is significant, making accurate detection and boundary localization difficult. To address these problems, this paper proposes an improved detection method based on YOLOv8n. In the baseline network, the C2f module is replaced with C2fHGBlockExp, the InterpIoU loss is adopted for bounding box regression, and a large separable kernel attention (LSKA) module is introduced after the multi-scale pooling features of SPPF to construct an LSKA-SPPF feature extraction module. Experiments are conducted on the public PV_Train_Val_28_12 dataset, which contains 2,781 images divided into 1,947 for training, 556 for validation, and 278 for testing, and all models are trained for 500 epochs. In the ablation study, the model combining the three improvements achieves Precision, Recall, mAP@0.5, and mAP@0.5:0.95 of 93.289%, 88.461%, 93.180%, and 69.438%, respectively, which are 9.890, 8.823, 7.889, and 11.653 percentage points higher than the baseline. The results show that the proposed combination of improvements enhances detection accuracy and localization performance under the given dataset and experimental settings.

Keywords

Boundary box regression, defect detection, electroluminescence imaging, large separable kernel attention, photovoltaic panels, YOLOv8n.

1. INTRODUCTION

Photovoltaic modules are exposed to outdoor environments for long periods, and defects such as cracks, micro-cracks, short-circuited cells, and other anomalies may appear in solar cells. These defects affect the electrical performance and operational reliability of the modules. Electroluminescence (EL) imaging can reveal internal structural differences of solar cells and provides an image basis for defect detection. As the scale of deployed modules continues to grow, manual inspection image by image can hardly balance efficiency and consistency; therefore, detection methods that can automatically locate and recognize multiple defect categories are required. Existing studies have used EL images to identify hidden cracks and other defects of photovoltaic modules, showing that EL imaging provides effective information for the visual detection of internal defects, and multi-source diagnosis combining thermal imaging with EL images also facilitates fault identification and health assessment of modules [1--3].

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202611_12(11).0007

YOLOv8n is a lightweight single-stage detection network with a compact model size, making it suitable as a baseline for resource-constrained detection tasks. However, in photovoltaic EL images, micro-cracks usually exhibit slender shapes and limited local contrast, and the scale and quantity distributions of different target categories are unbalanced; conventional feature extraction and bounding box regression may struggle to simultaneously account for detail expression, context modeling, and localization accuracy. Related reviews indicate that the YOLO series achieves a good balance between detection speed and accuracy, and that network structure, feature fusion, and loss functions are important directions for continuous improvement [7--8]. Studies on improving YOLOv8 for small targets and complex backgrounds also show that enhancing the use of shallow details and multi-scale feature expression helps improve detection performance [9--13].

To address the above problems, this paper makes three improvements to YOLOv8n: replacing the C2f feature module with C2fHGBlockExp, adopting the InterpIoU regression loss, and inserting an LSKA spatial modeling unit after the multi-scale pooling of SPPF to form an LSKA-SPPF module. Based on existing experimental records, this paper organizes the dataset, experimental settings, ablation results, and comparative results, with emphasis on the analysis of changes in each metric. Existing studies on EL defect detection of photovoltaic modules have carried out improvements from the perspectives of lightweight convolution, attention mechanisms, and non-maximum suppression, providing the application background and method reference for the module combination in this paper [1--4].

1.1. Related Work

Research on photovoltaic module defect detection has gradually expanded from traditional image processing to deep learning based object detection. Meng et al. constructed the YOLO-PV detection algorithm based on EL images, verifying the feasibility of a single-stage detection framework for EL defect recognition of photovoltaic modules. Domestic studies have further applied improved YOLOv4 to solar panel defect detection, and combined lightweight convolution, attention modules, and regression loss improvements within the YOLOv5 framework to enhance EL image detection [1--2]. These studies are directly related to the task of this paper; however, the category definitions, annotation rules, and dataset splits differ across datasets, and performance values should not be directly compared horizontally. As YOLOv8 has been increasingly applied, researchers have carried out targeted improvements on C2f feature extraction, multi-scale feature fusion, attention mechanisms, and bounding box regression. For example, existing studies have embedded EMA attention into the C2f module combined with the WIoU loss, and have adopted small-target detection layers, lightweight feature modules, or dynamic feature fusion structures to improve detection in complex backgrounds [3--5, 9--13]. In addition, surveys on object detection have systematically summarized the network structures and evaluation metrics of the YOLO series, and comparative studies on bounding box loss functions indicate that the choice of localization loss should be combined with target scale and data distribution [6--8, 14].

In summary, the existing literature provides the methodological basis for photovoltaic EL defect detection, YOLO network improvement, and bounding box regression optimization. However, the joint application of C2fHGBlockExp, LSKA-SPPF, and the InterpIoU implemented in the code of this study to the current six-category EL dataset still requires verification through ablation experiments under identical conditions. The experimental conclusions of this paper are limited to the dataset and training configuration used, and metrics reported on datasets of other publications are not treated as results of this study.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202611_12(11).0007

2. DATASET AND METHODS

2.1. Dataset and Split

The experiments adopt the PV_Train_Val_28_12 public object detection dataset published by the Faculty of Engineering, Fayoum University on Roboflow Universe [16]. The dataset contains 2,781 images with six categories: Examined, ShortCircuitString, ShortCircuitCell-LowPowerCell, Crack, MicroCrack, and OtherError. The split recorded in the experiments is 1,947 images for training, 556 for validation, and 278 for testing, totaling 2,781 images, with a ratio of about 7:2:1. The class distribution is listed in Table 1.

Table 1. Class distribution of the dataset

Class	Instance Count
Examined (normal/inspected)	1159
ShortCircuitString	420
ShortCircuitCell-LowPowerCell	995
Crack	431
MicroCrack	8978
OtherError	6837
Total	18820

2.2. YOLOv8n Baseline Network

YOLOv8n is a lightweight single-stage object detection network in the YOLO series, which performs target localization and category prediction simultaneously in a single forward pass [7--8]. The network consists of a backbone, a neck, and a detect head. The backbone gradually extracts image features and reduces the spatial resolution; the C2f module is responsible for feature transformation and fusion in the backbone stage, and the SPPF at the end enlarges the receptive field through successive pooling and aggregates multi-scale contextual information. The neck fuses features of different levels through top-down and bottom-up paths: deeper features are concatenated with shallow features after upsampling, and cross-scale information interaction is then performed through a downsampling path, so that semantic information and detail information jointly participate in detection. The detect head predicts bounding boxes and categories on feature layers of different resolutions (P3, P4, and P5); higher-resolution features help preserve small-target details, while lower-resolution features cover a larger context range. Figure 1 shows the overall pipeline of the baseline model from the input image, backbone feature extraction, and multi-scale feature fusion to the detection output. Feature layers of different scales participate in prediction in parallel, enabling the model to adapt to defect targets of different sizes.

2.3. Improved Network Structure

2.3.1. C2fHGBlockExp Backbone Feature Module

The C2f module in the baseline network is replaced with C2fHGBlockExp, forming the H-YOLOv8 model. The module inherits the outer structure of C2f: the input is first expanded by a 1×1 convolution to 2c channels and then split along the channel dimension into Y0 and Y1; Y0 is retained as a bypass feature, while Y1 passes through n HGBlockExp modules in series. The output of every HGBlockExp is retained, and finally Y0, Y1, and the outputs of all HGBlockExp stages are concatenated along the channel dimension and fused by a 1×1 convolution to c2 channels. The module pipeline is shown in Figure 2.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202611_12(11).0007

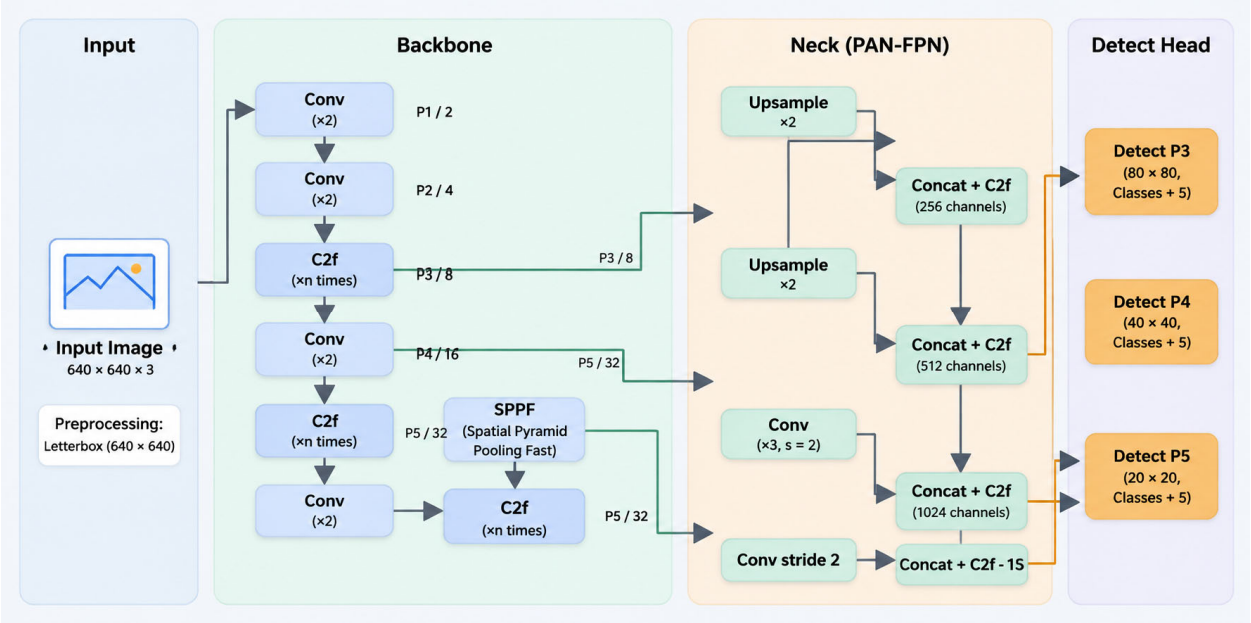


Figure 1. Overall structure of the YOLOv8n baseline network

Cracks and micro-cracks in EL images of photovoltaic panels usually appear in slender shapes, and some defects have low contrast, with local textures easily disturbed by the background structure of the cell. If shallow edge information gradually weakens during feature extraction, the model may struggle to use both defect details and a larger context simultaneously. C2fHGBlockExp retains a basic bypass feature and serially extracts features of multiple stages in another branch, then feeds the outputs of all stages together into the fusion layer, so that the raw branch and the progressively transformed features can be used in the same output. From a structural design perspective, this module helps enrich the hierarchical feature expression of tiny defects without additionally reducing the feature map resolution in the outer framework. In the ablation results, replacing C2f with C2fHGBlockExp alone improves mAP@0.5 and mAP@0.5:0.95 by 4.300 and 4.465 percentage points, respectively, verifying the effectiveness of the module under the experimental settings of this study. Related studies on photovoltaic EL images point out that defect regions may be disturbed by complex backgrounds, so the detection model needs to account for local texture, defect edges, and multi-scale feature information [1--5]; other small-target detection studies have also enhanced shallow features and improved C2f-based feature extraction structures to improve target representation [9--13].



Figure 2. Improved structure of C2fHGBlockExp

Structure formulas and execution procedure.

Let the input feature be X ∈ R^(c1×H×W), the expansion ratio be e, and the hidden channel number be c = [e·c2]. According to the given implementation of C2fHGBlockExp, its outer computation can be expressed as follows:

$$c = [\, e \cdot c2 \,] \tag{1}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202611_12(11).0007

$$Conv1\times1(X) = SiLU(BN(W1\times1 * X)) \quad (2)$$

$$U = Conv1\times1(X), \quad U \in R^{(2c\times H\times W)} \quad (3)$$

$$(Y0, Y1) = Split(U, c), \quad Y0, Y1 \in R^{(c\times H\times W)} \quad (4)$$

$$H1 = Y1; \quad H(i+1) = HGBlockExp_i(Hi), \quad i = 1, ..., n \quad (5)$$

$$V = Concat(Y0, H1, H2, ..., H(n+1)), \quad V \in R^{((n+2)c\times H\times W)} \quad (6)$$

$$Z = Conv1\times1(V), \quad Z \in R^{(c2\times H\times W)} \quad (7)$$

The execution procedure of each step is as follows. (1) Channel expansion: cv1 applies a 1×1 convolution, batch normalization, and SiLU activation to the input feature, changing the number of channels to 2c. (2) Channel split: the output of cv1 is equally divided along the channel dimension into Y0 and Y1, each containing c channels, and Y0 is retained as the bypass feature. (3) Serial feature extraction: Y1 passes through n HGBlockExp modules in series; each module receives the output of the previous one, forming H2 to H(n+1), and the output of every stage is retained. (4) Feature concatenation: Y0, Y1, and the n HGBlockExp outputs are concatenated along the channel dimension, giving (2+n)c channels. (5) Channel fusion: cv2 applies another 1×1 convolution, batch normalization, and SiLU activation to fuse the concatenated features into c2 channels, and the output spatial size remains H×W.

Internal structure of HGBlockExp.

HGBlockExp first retains the input feature and then extracts features layer by layer through m Conv or LightConv blocks. The output of every level participates in concatenation; the concatenation result passes sequentially through an SC channel-compression convolution and an EC channel-mapping convolution. If the shortcut is enabled and the input and output channel numbers are identical, the module finally adds the input to the mapped result; otherwise it directly outputs the mapped result. The structure is shown in Figure 3.

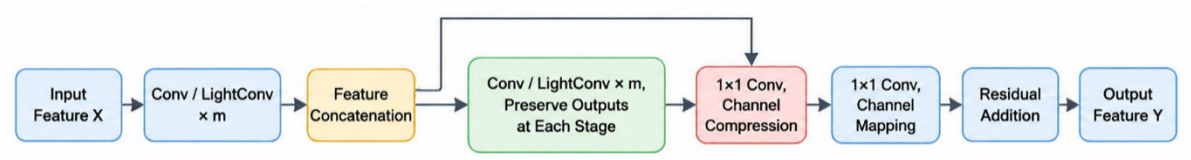


Figure 3. Internal structure of HGBlockExp

Let the input be X, the internal repetition number be m, and the intermediate channel number be cm. The first convolution block maps c1 channels to cm, and the subsequent convolution blocks process the features level by level within cm channels. The computation can be written as:

$$Q0 = X; \quad Qi = Bi(Q(i-1)), \quad i = 1, ..., m \quad (8)$$

$$F = Concat(X, Q1, Q2, ..., Qm), \quad F \in R^{((c1+m\cdot cm)\times H\times W)} \quad (9)$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202611_12(11).0007

$$S = \text{Conv1×1_SC}(F), \quad S \in R^{((c2//2)×H×W)} \tag{10}$$

$$E = \text{Conv1×1_EC}(S), \quad E \in R^{(c2×H×W)} \tag{11}$$

$$Y = E + X, \text{ if shortcut = True and } c1 = c2; \text{ otherwise } Y = E \tag{12}$$

When lightconv = False, Bi is an ordinary k×k convolution block executed in the order Conv2d, BatchNorm, and ReLU; when lightconv = True, Bi first adjusts the channels through a 1×1 convolution without activation, and then goes through a depthwise convolution, BatchNorm, and the activation function. The autopad pads according to the kernel size when the stride is 1, keeping the spatial size of the feature map unchanged:

$$\text{ConvBlock}(x) = \text{ReLU}(\text{BN}(\text{Convk×k}(x))) \tag{13}$$

$$\text{LightConv}(x) = \text{ReLU}(\text{BN}(\text{DWConvk×k}(\text{BN}(\text{Conv1×1}(x))))) \tag{14}$$

In the C2fHGBlockExp example, the parameters of each HGBlockExp are c1 = c2 = cm = 128, m = hg_n = 1, k = 3, lightconv = False, and shortcut = False. After the input and the convolution output of this stage are concatenated to 256 channels, SC compresses them to 64 channels and EC maps them back to 128 channels; since the shortcut is disabled, the residual addition is not performed in this example. If the module independently adopts m = 6, the number of concatenated channels is computed as c1 + 6·cm.

2.3.2. InterpIoU Bounding Box Regression Loss

The bounding box regression loss measures the localization difference between the predicted box and the ground-truth box. This paper adopts the InterpIoU loss (its definition and implementation follow the training code records of this study). According to the provided training code, the interpolation coefficient is set to $\alpha = 0.98$; the method constructs an interpolation box between the predicted box and the ground-truth box, and jointly optimizes the IoU loss between the two boxes and the IoU loss between the interpolation box and the ground-truth box. Let the predicted box be B_pred and the ground-truth box be B_gt:

$$B_{int} = (1 - 0.98)\cdot B_{pred} + 0.98\cdot B_{gt} \tag{15}$$

$$L_{InterpIoU} = L_{IoU}(B_{pred}, B_{gt}) + L_{IoU}(B_{int}, B_{gt}) \tag{16}$$

$$L_{IoU}(A, B) = 1 - IoU(A, B) \tag{17}$$

where B_int is the interpolation box, and $\alpha = 0.98$ keeps the interpolation box close to the ground-truth box. The second loss term provides an additional optimization signal for the interpolation box; when the predicted box and the ground-truth box do not overlap, the interpolation box may still intersect the ground-truth box, thereby supplementing the gradient of the standard IoU loss. Replacing the original CIoU term with the above InterpIoU loss means that the localization objective becomes the IoU loss between the predicted box and the ground-truth box plus the IoU loss between the interpolation box and the ground-truth box.

2.3.3. LSKA-SPPF Multi-Scale Spatial Feature Module

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202611_12(11).0007

The LSKA-SPPF module adds a large separable kernel attention after the serial pooling structure of SPPF to fuse context information of different receptive fields and enhance spatial feature expression. The input feature X is first compressed by a 1×1 convolution, followed by three successive max-pooling operations; the initial compressed feature and the pooling results of all levels are concatenated and then fed into LSKA for spatial feature modeling, and a final 1×1 convolution completes the channel fusion to produce the output feature Z. This structure retains the characteristic of SPPF of gradually enlarging the receptive field and supplements long-range spatial dependency modeling through LSKA. In YOLOv8 improvement studies, embedding attention mechanisms into the C2f module or the feature fusion path has been used to enhance multi-scale target feature expression; the related results provide the design reference for adding a spatial modeling unit after the pooled features in this paper [9--13]. The module structure is shown in Figure 4.

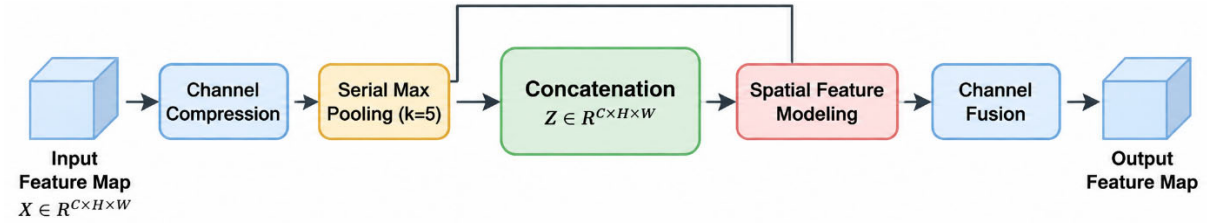


Figure 4. Structure of the LSKA-SPPF multi-scale spatial feature module

3. EVALUATION METRICS

Precision, Recall, F1 score, mAP@0.5, and mAP@0.5:0.95 are used to evaluate the detection performance. Precision and Recall measure the proportion of correct detections among the targets predicted as positive and the proportion of real targets successfully detected, respectively. The F1 score is the harmonic mean of Precision and Recall and is used to comprehensively evaluate the precision-recall balance; the higher its value, the better the combined performance in reducing false positives and missed detections. These metrics are commonly used performance measures in object detection research, and in particular the variation of mAP under different IoU thresholds helps analyze classification and localization performance [7--8]:

$$\text{Precision} = TP / (TP + FP) \tag{18}$$

$$\text{Recall} = TP / (TP + FN) \tag{19}$$

$$F1 = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall}) \tag{20}$$

where TP denotes true positives, i.e., the number of targets correctly detected by the model; FP denotes false positives, i.e., the number of background or non-target regions incorrectly recognized as targets; and FN denotes false negatives, i.e., the number of targets that exist but are not detected. The F1 score ranges from 0 to 1, and a value closer to 1 indicates better combined precision and recall performance. Unlike evaluating the model with Precision or Recall alone, the F1 score considers the effects of both false positives and missed detections and can therefore serve as an auxiliary measure of the overall detection performance.

AP is the area under the precision-recall curve, and mAP is the average of AP over all categories. mAP@0.5 computes the AP of each category at an IoU threshold of 0.5 and averages

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202611_12(11).0007

them; mAP@0.5:0.95 computes AP at IoU thresholds from 0.50 to 0.95 in steps of 0.05 and averages the results. Compared with mAP@0.5 at a single IoU threshold, mAP@0.5:0.95 imposes stricter requirements on the localization accuracy of predicted bounding boxes and reflects the object detection performance of the model more comprehensively.

4. EXPERIMENTAL SETUP

The experimental platform and training settings are summarized in Table 2 according to the original records. To determine the number of training epochs, a 1,000-epoch training run was first conducted to observe the metric variations; the records show that after about 500 epochs, Precision, Recall, and mAP tend to stabilize with limited further improvement. Therefore, all subsequent comparative and ablation experiments uniformly use 500 training epochs.

Table 2. Experimental environment and training configuration

Item	Configuration
GPU	NVIDIA RTX 5060
CPU	AMD Ryzen 5 7500F
Operating system	Windows 10
Deep learning framework	PyTorch 2.11.0
GPU acceleration	CUDA 13.3
Programming language	Python 3.12.7
Training epochs	500

YOLOv8n is used as the baseline detection network and is iteratively optimized with the training set. During training, the input image size is set to 640 × 640 pixels, the batch size to 16, and the number of epochs to 500, with 4 data-loading worker processes. The momentum coefficient is set to 0.937 and the weight decay coefficient to 0.0005. In addition, the warm-up period is set to 3 epochs, and automatic mixed precision training is enabled to improve the computational efficiency. The random seed is set to 0, and deterministic training is enabled to enhance the reproducibility of the experiments. The hyperparameter settings are listed in Table 3.

Table 3. Training hyperparameter settings

Hyperparameter	Setting
Image size	640 × 640
Workers	4
Epochs	500
Batch size	16
Weight decay	0.0005
Momentum	0.937

5. RESULTS AND ANALYSIS

5.1. Comparative Experiment

According to the updated experimental results, Table 4 lists the comparison between the YOLOv8n baseline and the improved model.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202611_12(11).0007

Table 4. Comparison between YOLOv8n and the improved model

Model	P (%)	R (%)	F1 Score	mAP@0.5 (%)	mAP@0.5:0.95 (%)
YOLOv8n	83.399	79.638	81	85.291	57.785
HL-YOLOv8	93.289	88.461	91	93.18	69.438

Compared with YOLOv8n, the improved model improves P, R, mAP@0.5, and mAP@0.5:0.95 by 9.890, 8.823, 7.889, and 11.653 percentage points, respectively.

5.2. Confusion Matrix Analysis

Figures 5 and 6 show the normalized confusion matrices of the baseline model and the improved model, respectively. On the six categories Examined, ShortCircuitString, ShortCircuitCell-LowPowerCell, Crack, MicroCrack, and OtherError, the diagonal values of the improved model are 1.00, 0.95, 0.92, 0.86, 0.91, and 0.81 in sequence, while the corresponding values of the baseline model are 1.00, 0.93, 0.86, 0.70, 0.83, and 0.54. Except for Examined, the diagonal classification ratios of all categories increase, among which OtherError improves by 0.27 and Crack by 0.16, indicating that the improved model ameliorates the recognition of these easily confused defect categories. The background-related missed detection ratio also decreases significantly; for example, the background row value corresponding to OtherError drops from 0.45 to 0.19.

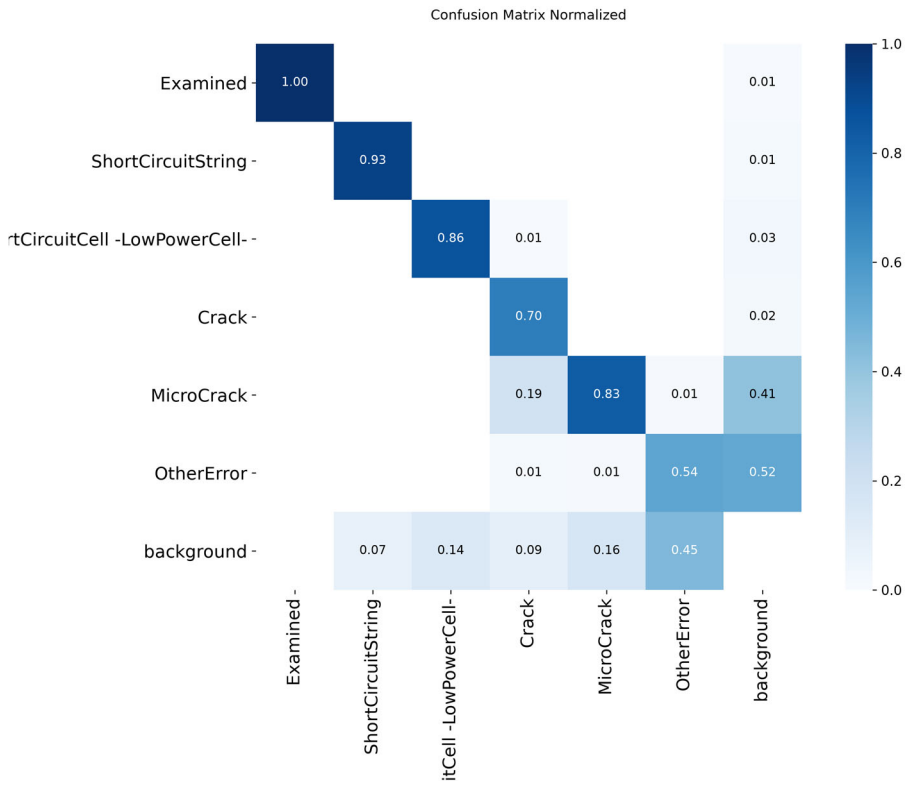


Figure 5. Normalized confusion matrix of the YOLOv8n baseline model

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202611_12(11).0007

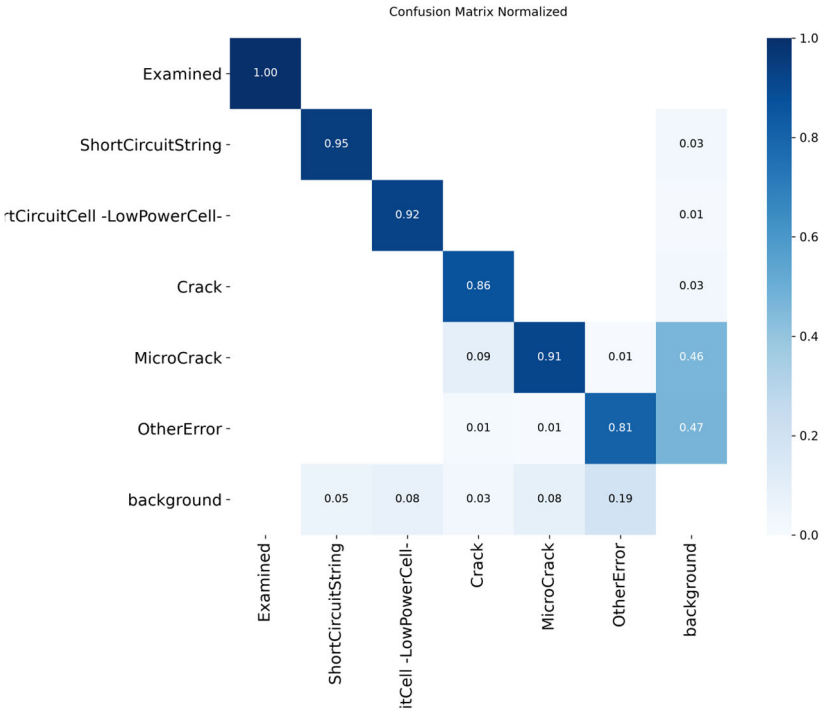


Figure 6. Normalized confusion matrix of the improved model

5.3. Training Metric Curves

Based on the two results.csv files, the curves of Precision, Recall, mAP@0.5, and mAP@0.5:0.95 versus training epoch are plotted in Figure 7. Both models are trained for 500 epochs, and the four metrics in the final epoch are consistent with the records in Table 4. The improved model exhibits higher validation metrics throughout the training process, reaching P, R, mAP@0.5, and mAP@0.5:0.95 of 93.289%, 88.461%, 93.180%, and 69.438% at the final epoch, respectively, compared with 83.399%, 79.638%, 85.291%, and 57.785% for the baseline. The curves show how the metrics vary with the training progress and can be used to observe the convergence trend; they reflect the records of these training runs and do not represent the statistical fluctuation of repeated experiments.

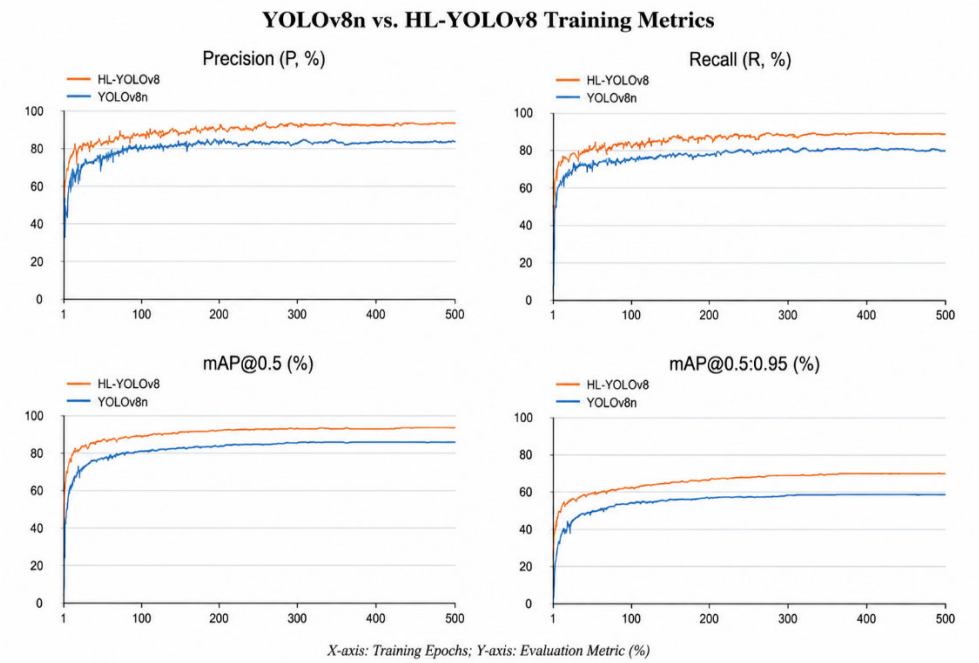


Figure 7. Training metric curves of YOLOv8n and HL-YOLOv8

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202611_12(11).0007

5.4. Validation Set Detection Visualization

Figure 8 shows the visualization of the ground-truth labels of a validation batch of HL-YOLOv8, and Figure 9 shows the corresponding prediction results. Targets are marked with boxes of different colors and category names, and the prediction map also shows the confidence of each detected box, which can be used to visually examine the localization and recognition of the model for cell regions and different defect categories. As the mosaic contains a large number of targets and dense annotation boxes, some local text and bounding boxes overlap; this visualization demonstrates the batch-level detection effect, and specific false positives and missed detections should still be verified item by item against the original images and annotations.



Figure 8. Visualization of ground-truth labels for a validation batch of HL-YOLOv8



Figure 9. Visualization of the corresponding prediction results for the validation batch

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202611_12(11).0007

5.5. Ablation Experiment

To analyze the roles of the backbone module, InterpIoU, and LSKA-SPPF, the strategies are combined under both the C2f and C2fHGB backbone settings, and the results are listed in Table 5. The symbol "√" indicates that the corresponding improvement is adopted. Existing studies on photovoltaic EL defect detection have carried out ablation or comparative verification by means of module replacement, attention enhancement, and regression loss adjustment [1--5]; therefore, this paper combines the three strategies under unified data splits and training epochs.

Table 5. Ablation results of different improvement combinations

Backbone	LSKA-SPPF	InterpIoU	P (%)	R (%)	F1 Score	mAP@0.5 (%)	mAP@0.5:0.95 (%)
C2f	—	—	83.399	79.638	81	85.291	57.785
C2f	√	—	84.903	80.695	82	86.005	58.368
C2f	—	√	87.37	82.791	85	88.076	60.776
C2f	√	√	89.204	84.247	86	89.843	62.67
C2fHGB	—	—	89.879	85.356	87	90.577	64.454
C2fHGB	√	—	90.305	86.879	88	91.329	64.777
C2fHGB	—	√	92.475	87.741	90	93.03	67.399
C2fHGB	√	√	93.289	88.461	91	93.18	69.438

Figure 10 shows the variation of Precision, Recall, mAP@0.5, and mAP@0.5:0.95 over the 500 training epochs for the eight ablation configurations. All four metrics gradually increase with the training progress and tend to stabilize in the late stage; clear performance hierarchies exist among the curves of different improvement combinations, and the complete configuration adopting C2fHGB, LSKA-SPPF, and InterpIoU together generally maintains the lead in the late training stage. Taking the C2f baseline as the reference, the complete configuration achieves P, R, mAP@0.5, and mAP@0.5:0.95 of 93.289%, 88.461%, 93.180%, and 69.438% at epoch 500, which are 9.890, 8.823, 7.889, and 11.653 percentage points higher than the reference, respectively.

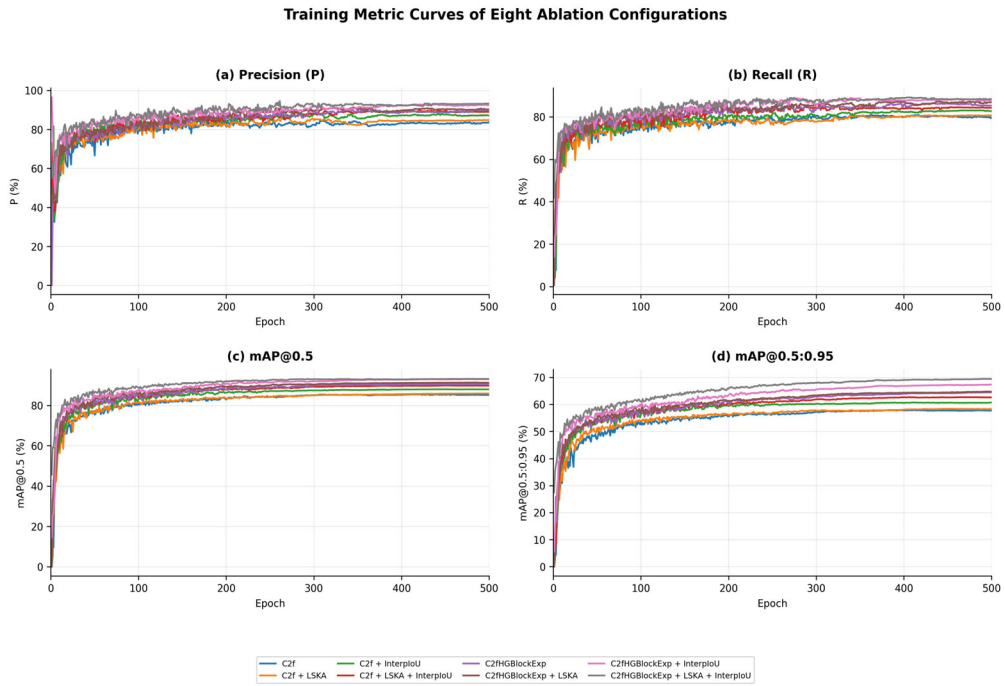


Figure 10. Training metric curves of the eight ablation configurations

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202611_12(11).0007

6. CONCLUSION

Aiming at the problems of photovoltaic panel electroluminescence (EL) images, such as tiny and slender defect features, large scale differences among categories, inconspicuous local textures, and difficult bounding box localization, this paper proposes an improved detection method for surface defects of photovoltaic panels based on YOLOv8n. Taking YOLOv8n as the baseline network, the method is improved in three aspects: backbone feature extraction, bounding box regression optimization, and multi-scale spatial feature modeling. First, the C2fHGBlockExp module replaces the original C2f module to enhance the extraction and fusion of features at different levels. Second, the InterpIoU loss is introduced; by constructing an interpolation box between the predicted box and the ground-truth box, it provides an additional optimization constraint for bounding box regression. Finally, a large separable kernel attention (LSKA) is added after the multi-scale pooled features of the SPPF module to construct the LSKA-SPPF module, enhancing the modeling of spatial features and information of different receptive fields. Through these improvements, the HL-YOLOv8 detection model integrating multiple feature enhancement strategies is formed.

To verify the effectiveness of each improvement, comparative and ablation experiments are conducted on the PV_Train_Val_28_12 dataset. The results show that each improvement alone improves the detection performance to a certain extent, and their combination further enhances the overall detection effect, with the joint use of C2fHGBlockExp, InterpIoU, and LSKA-SPPF achieving the best overall performance. The final HL-YOLOv8 model reaches Precision, Recall, mAP@0.5, and mAP@0.5:0.95 of 93.289%, 88.461%, 93.180%, and 69.438%, respectively, which are 9.890, 8.823, 7.889, and 11.653 percentage points higher than the original YOLOv8n baseline. In particular, the mAP@0.5:0.95 metric improves by 11.653 percentage points, indicating that the improved model has better comprehensive detection performance under multiple IoU thresholds and that the bounding box localization capability is also improved.

From the training curves of the eight ablation configurations, the validation metrics of all configurations generally increase and then stabilize with the training progress. Compared with the original C2f baseline configuration, the complete model adopting C2fHGBlockExp, InterpIoU, and LSKA-SPPF maintains better metric performance in the late training stage, indicating that the three improvements are complementary under the experimental conditions of this study. Specifically, C2fHGBlockExp improves feature expression from the perspective of the feature extraction structure, InterpIoU focuses on optimizing the regression relationship between the predicted box and the ground-truth box, and LSKA-SPPF further strengthens the spatial information modeling of multi-scale features. Acting together from different stages, the three enable the model to make fuller use of defect features in EL images and to improve the detection of different types of surface defects of photovoltaic panels. The confusion matrix results also show that the classification performance of the improved model improves to varying degrees on categories such as short-circuit string, short-circuit cell, crack, micro-crack, and other defects, indicating that the method has good recognition effect on some easily confused defect categories.

REFERENCES

[1] Yaermaimaiti, Y., Deng, H., & Xie, L. (2023). Solar panel defect detection based on improved YOLOv4. Journal of South China Normal University (Natural Science Edition), 55(5), 21–30. (in Chinese)

[2] Zheng, Z., & Ma, P. (2025). EL image defect detection of photovoltaic modules based on improved YOLOv5. Acta Energiae Solaris Sinica, 46(10), 295–301. (in Chinese)

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202611_12(11).0007

[3] Tang, Z., & Tang, D. (2025). Solar panel defect detection based on improved YOLOv8. Journal of Hunan University of Science and Technology (Natural Science Edition), (4). (in Chinese)

[4] Wang, T., & Li, Y. (2026). Photovoltaic cell defect detection algorithm based on improved YOLOv8n. Journal of Chongqing Technology and Business University (Natural Science Edition), 43(4), 42–51. (in Chinese)

[5] Chen, T., & Chen, Z. (2026). EL image based photovoltaic cell defect detection model for edge devices. Journal of Fuzhou University (Natural Science Edition), 54(1), 1–9. (in Chinese)

[6] Liu, X., & Ge, M. (2025). Fault diagnosis and health assessment of photovoltaic modules based on image processing technology. Solar Energy, (12), 51–58. (in Chinese)

[7] Xu, Y., & Li, J. (2024). A survey of YOLO series object detection algorithms. Computer Science and Exploration, 18(9), 2221–2238. (in Chinese)

[8] Shao, Y., & Zhang, D. (2022). A survey of deep learning based YOLO object detection. Journal of Electronics & Information Technology, 44(10), 3697–3708. (in Chinese)

[9] Wang, X., Chang, J., & Yu, Q. (2025). Research on small object detection methods based on YOLOv8. Journal of North University of China (Natural Science Edition), 46(6), 712–725. (in Chinese)

[10] Liu, J., & Shi, R. (2025). An improved YOLOv8-based small target detection model for UAV vision. Journal of Northeastern University (Natural Science), 46(12), 29–37. (in Chinese)

[11] Cao, J., & Qiao, G. (2025). YOLO algorithm improvement strategies for small object detection from high-altitude perspectives. Journal of Computer Applications. (in Chinese)

[12] Cheng, H., & Jiao, L. (2024). Improved YOLOv8 for remote sensing image detection. Radio Engineering, 54(5), 1155–1161. (in Chinese)

[13] Cheng, H., & Qiao, Q. (2024). Object detection in UAV aerial images based on improved YOLOv8. Radio Engineering, 54(4), 871–881. (in Chinese)

[14] Dong, H., & Pan, J. (2024). Research on bounding box regression loss functions based on the YOLOv5s model. Modern Electronic Technology, 47(3), 179–186. (in Chinese)

[15] Huo, J., & Su, H. (2025). Small target vehicle detection in road traffic based on improved YOLOv8. Computer Engineering, 51(1), 246–257. (in Chinese)

[16] Faculty of Engineering Fayoum University. (2022). PV_Train_Val_28_12 Dataset. Roboflow Universe. https://universe.roboflow.com/faculty-of-engineering-fayoum-university/pv_train_val_28_12